III. METHODS FOR ANALYZING CANCER INCIDENCE DATA

A. Case Identification

The observed number of cancer cases in this evaluation was derived from cases reported to the Massachusetts Cancer Registry (MCR) as primary site cancer cases diagnosed in Wayland residents during 1982 through 1995. Cases were selected for inclusion based on the address reported to the hospital or reporting facility at the time of diagnosis.

The MCR began collecting information on Massachusetts residents diagnosed with cancer in the state in 1982. All newly diagnosed cancer cases are required by law to be reported to the MCR within six months of the date of diagnosis (M.G.L.c.111s.111B). The 14-year period 1982-1995 constitutes the period for which the most recent and complete cancer incidence data were available at the time of this analysis.

The term "cancer" is used to describe a variety of diseases associated with abnormal cell and tissue growth. Primary site (location in the body where the disease originated) and histology (tissue or cell type) classify the different cancer types. Epidemiological studies have revealed that different types of cancer are individual diseases with separate causes, risk factors, characteristics and patterns of survival (Bang 1996).

Ten types of cancer were evaluated in this investigation. These include cancers of the bladder, breast, kidney, liver, lung (including bronchus), pancreas, and stomach, as well as Hodgkin's disease, non-Hodgkin's lymphoma (NHL) and leukemia. These cancer types were selected in order to address concerns expressed by residents of Wayland regarding suspected elevations in the incidence of these cancer types.

Only primary site cancers were included in this evaluation. Cancers that occur as the result of the metastases or the spread of a primary site cancer to another location in the body are not considered as a separate cancer and therefore, were not included. It should be noted that in 1992 the MCR began collecting data on cancers categorized as in situ cancers. In situ

cancers are malignant cells in the earliest stages of development, which do not always progress to invasive cancers. Since the public comment release, the BEHA no longer includes in situ cancers in cancer incidence evaluations because these data are not complete and are only available since 1992. Therefore, although in situ cancers were included in the previous evaluation (1982-1992), they have not been included in the analysis of more recent cancer incidence data (1987-1994, 1995). It should also be noted that the omission of in situ cancers only affects one cancer type, bladder cancer, resulting in one less observed case for the time period 1987-1994.

Occasionally, the MCR research file may contain duplicate cases. The data discussed in this report have been controlled for duplicate cases by excluding them from the analyses. However, reports of multiple primary site cancer cases were included. Duplicate cases are additional reports of the same primary site cancer case. A multiple primary cancer case is defined by the MCR as a new tumor of the same histology (original location in the body) more than two months after the initial diagnosis (MCR 1996). The determination that a case was a duplicate and should be excluded from the

analyses was made by the MCR after consulting with the hospital or reporting facilities and obtaining additional information regarding the histology and/or pathology of the case.

B. Calculation of Standardized Incidence Ratios (SIRs)

To determine whether elevated numbers of cancer cases have occurred in Wayland or its census tracts, cancer incidence data were analyzed by age and gender to compare the observed number of cancer cases in each census tract to the number that would be expected based on the statewide cancer experience. Standardized Incidence Ratios (SIRs) were calculated for the period 1982-1994 for each of the ten cancer types for the two CTs and the town as a whole. SIRs were also calculated for the three time periods 1982-1986, 1987-1992 and 1987-1994 in order to evaluate temporal trends in cancer incidence.

An SIR is an estimate of the occurrence of disease in a population in relation to what might be expected if the population in question had the same cancer experience as some

larger population designated as the comparison population. Usually, the state as a whole is selected to be the "comparison" population. Using the state of Massachusetts as a comparison population provides a stable population base for the calculation of incidence rates. As a result of the instability of incidence rates based on small numbers of cases, SIRs were not calculated when fewer than five cases were observed.

Specifically, an SIR is the ratio of the observed number of cancer cases to the expected number of cases multiplied by 100. An SIR of 100 indicates that the number of cancer cases observed in the population evaluated is equal to the number of cancer cases expected in the comparison or "normal" population. An SIR greater than 100 indicates that more cancer cases occurred than expected and an SIR less than 100 indicates that fewer cancer cases occurred than expected. Accordingly, an SIR of 150 is interpreted as 50% more cases than the expected number; an SIR of 90 indicates 10% fewer cases than expected.

Caution should be exercised, however, when interpreting an SIR. The interpretation of an SIR depends on both the size and stability of the SIR itself. Two SIRs can have the same size but not
have the same stability. An SIR of 150 based on six observed cases and four expected cases indicates a 50% excess in cancer, but this excess is actually only two cases. Conversely, an SIR of 150 based on 600 observed cases and 400 expected cases represents the same 50% excess in cancer, but because the SIR is based upon a greater number of cases, the estimate is more stable. It is unlikely that 200 excess cases of cancer would occur by chance alone.

In order to calculate incidence rates, it is necessary to obtain accurate population information. The population figures used in this analysis were interpolated based on 1980 and 1990 census data for Wayland (US DOC 1980, 1990). Midpoint estimates were calculated for each time period evaluated. To estimate the population between census years, an assumption was made that the change in population occurred at a constant rate throughout the ten-year interval between each census.

C. Calculation of 95% Confidence Interval

In addition to calculating SIRs, the statistical significance of each SIR was also assessed. A 95% confidence interval (95% CI) was calculated for each SIR to determine if the observed number of cases is significantly different from the expected number or if the difference may be due solely to chance (Rothman and Boice 1982). A 95% CI is a method of assessing the magnitude and stability of an SIR. Specifically, a 95% CI is the range of estimated SIR values that have a 95% probability of including the true SIR for the population. If the 95% CI range does not include the value 100, then the study population is significantly different from the comparison or "normal" population. "Significantly different" means there is less than a 5% chance that the observed difference is the result of random fluctuation in the number of observed cancer cases.

For example, if a confidence interval does not include 100 and the interval is above 100 (e.g., 105-130), then this means statistically there is a significant excess in the number of cancer cases. Similarly, if the confidence interval does not include 100 and the interval is below 100 (e.g., 45-96), then statistically the number of cancer cases is significantly lower than expected. If the confidence interval range includes 100, then the true SIR may be 100, and it cannot be concluded with sufficient confidence that the observed number of cases is not a result of chance and reflects
a real cancer increase or decrease. Again, as a result of the instability of incidence rates based on small numbers of cases, statistical significance was not assessed when fewer than five cases were observed.

In addition to the range of the SIR estimates contained in the confidence interval, the width of the confidence interval also reflects the stability of the SIR estimate. For example, a narrow confidence interval (e.g., 103-115) allows a fair level of certainty that the calculated SIR is close to the true SIR for the population. A wide interval (e.g., 85-450) leaves considerable doubt about the true SIR, which could be much lower than or much higher than the calculated SIR. This would indicate an unstable statistic.

D. Determination of Geographic Distribution

In Wayland, the geographic distribution of cancer was determined using available address information from the MCR indicating residence at the time of diagnosis. This information was mapped for each individual using a computerized geographic information system (GIS) (MapInfo 1996). This allowed the assignment of census tract location for each case as well as an evaluation of the spatial distribution of cases at a smaller geographic level (i.e., neighborhoods). The geographic distribution was assessed using a qualitative evaluation of the point pattern of cases within the town and within each census tract. In instances where the address information was incomplete (i.e., did not include specific streets or street numbers), efforts were made to research those cases using telephone books and town residential lists issued within two years of an individual's diagnosis. Address locations were also confirmed by site visits to the area.

E. Evaluation of Cancer Risk Factors

The MCR routinely collects data related to possible risk factors for individual cancer cases (e.g., smoking status, and occupation). Smoking status information was reviewed for cancers with known or suggested associations with tobacco smoke and occupational information was reviewed for cancer types that have been associated with exposures in specific occupations. In addition,
available breast cancer case information for the years 1987-1994 was reviewed to evaluate the stage of the cancer at time of diagnosis.

However, information about personal risk factors (e.g., family history, hormonal events, diet) that may also influence the development of cancer is not collected by the MCR and, therefore, could not be evaluated in this investigation. In addition, many cancers have a lengthy latency period. The latency period is the interval between first exposure to a disease-causing agent(s) and the appearance of symptoms of the disease (Last 1995).

Cancer does not usually develop within months after exposure. For most cancers, the latency period is an interval usually between 12 to 25 years and in some cases may be more than 40 or 50 years (Bang 1996, Frumkin 1995).